LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-CONF-661659

Incremental SVD for Insight into Wind Generation

C. Kamath, Y. J. Fan

September 30, 2014

The 13th International Conference on Machine Learning and Applications (ICMLA'14)
Detroit, MI, United States
December 3, 2014 through December 6, 2014

Incremental SVD for Insight into Wind Generation

Chandrika Kamath and Ya Ju Fan
Center for Applied Scientific Computing
Lawrence Livermore National Laboratory
Livermore, CA
Email: kamath2, fan4@llnl.gov

***Abstract*—In this paper, we formulate the problem of predicting wind generation as one of streaming data analysis. We want to understand if it is possible to use the weather data in a time window just before the current time to gain insight into how the wind generation might behave in a time interval just after the current time. Specifically, we use a singular value decomposition of the weather data, and show that the number of singular values and the largest singular value can be used to predict the magnitude of the change in the generation in the near future. The analysis uses an incremental algorithm based on a sliding window for reduced computational costs.**

I. INTRODUCTION

Integrating wind energy on the power grid is an extremely challenging problem. Wind is an intermittent resource and is typically scheduled using a forecast. These forecasts are obtained using numerical weather prediction techniques or derived from historical data. In our previous work, we considered ways in which we can use data mining techniques to provide control room operators additional information they can use while scheduling wind resources. In [1], we identified diurnal motifs, or recurring patterns, in the data. We found that there are a limited number of these motifs, and if we know the shape of the generation during the early hours of the day, we can predict how the generation is likely to change later in the day. In other related work [2], we used weather data and feature selection techniques to identify important variables relevant to days with ramp events, where the generation changes by a large amount in a short time. By monitoring just the important variables, the operators could reduce their data overload. We also showed that classification techniques could be helpful in predicting days when ramp events were likely to occur.

In the current work, we take a slightly different approach to the use of data mining in scheduling wind energy. We view the problem as one of streaming data analysis and investigate what we might be able to learn about the wind generation in the near future based on the weather conditions in the time window just before the current time. In contrast to our earlier work, where the temporal resolution used was a day, we expect that by using a finer temporal resolution of 1-3 hours in our analysis, we can capture the short term changes in the weather that might influence the generation in the near future.

This paper is organized as follows: First, in Section II, we describe the wind energy generation and weather data from the mid-Columbia Basin that are used in our analysis. In Section III, we formulate our problem as one of transforming the large number of weather variables in a time window just before the current time into a lower dimensional space. We describe how we can use the characteristics of this lower dimensional space to understand the behavior of the wind generation in the time period just after the current time. In Section IV, we describe an incremental algorithm that updates the lower dimensional space as new data arrive and the weather variables corresponding to the oldest time stamp are removed from the time window. Section V describes the insight we gain from our analysis, Section VI discusses related work and Section VII concludes with a summary.

II. DESCRIPTION OF THE DATA

The wind energy generation data used in our work are obtained from Bonneville Power Administration (BPA) in the mid-Columbia Basin region. We use the total generation from all wind farms in the BPA balancing area [3]. The data are sampled at 5-minute intervals. Figure 1 shows the generation for the first week in October 2011.

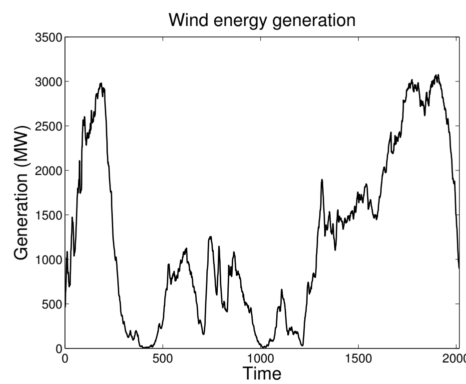


Fig. 1. The total wind generation for the first week in October 2011.

The weather data used in our analysis are from weather stations near the BPA balancing area. Some of these stations have data missing over long time periods. So we chose eight stations that have complete data. The names of these stations and the seven variables available at each location are listed in Table I. Figure 2 shows the variables at Biddle Butte for the first week in October 2011. Like the wind generation data, the weather data are also sampled at 5 minute intervals.

Weather Station	Measurement
Augspurger (A)	barometric pressure
Biddle Butte (B)	relative humidity
Hood River (HR)	temperature
Mt Hebo (MH)	wind direction
Roosevelt (R)	wind speed
Sunnyside (SS)	peak speed
Tillamook (Ti)	peak direction
Troutdale (Tr)	

TABLE I. THE EIGHT WEATHER STATIONS NEAR THE BPA BALANCING AREA. EACH STATION HAS SEVEN WEATHER MEASUREMENTS, RESULTING IN 56 VARIABLES.

Both the wind energy generation data and the weather data contain some missing time periods. We use interpolation to fill in the missing values in short time periods. If the missing values are in a longer time period, we remove the entire time segment from our analysis.

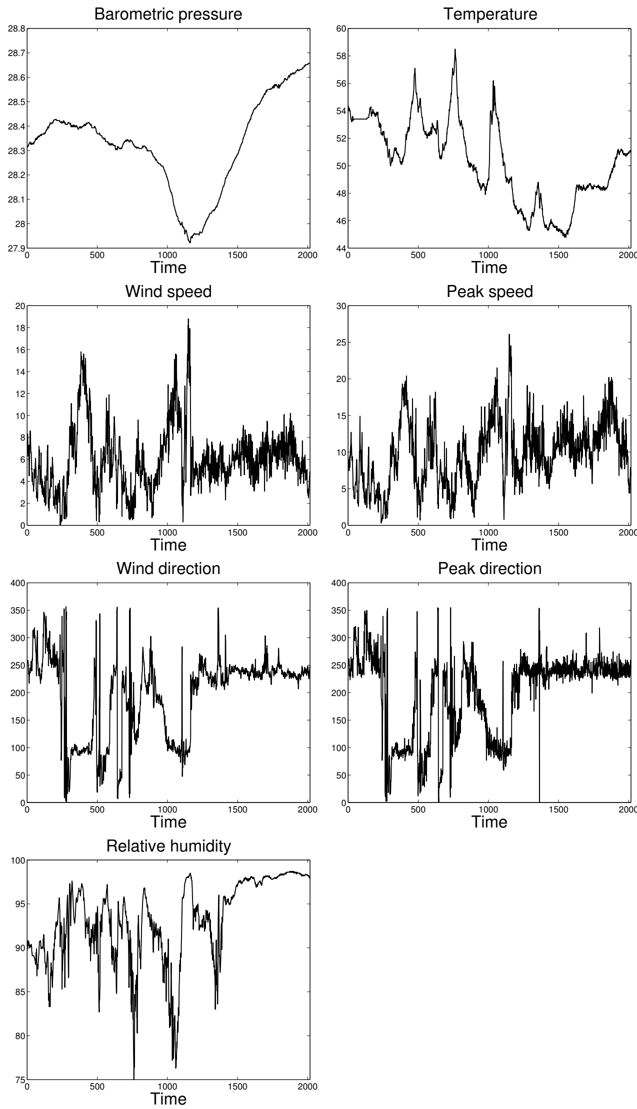


Fig. 2. The seven weather variables - barometric pressure, temperature, wind speed, peak wind speed, wind direction, peak wind direction, and relative humidity - for the first week in October 2011 at the Biddle Butte station.

III. PROBLEM FORMULATION

The question we want to address is the following: can we exploit the weather data in a time window of length c just before the current time t, to gain some insight on the generation from time period t to $t+l$? There are many ways in which we can address this question depending on what insight we want to gain about the generation and how we process the data in the time window. For our dataset, since the number of variables is large, we consider a lower-dimensional representation of the data. We expect that this representation will change over time and investigate if the characteristics of the representation are correlated with the generation in the time period just after the current time.

We consider the singular value decomposition (SVD) to obtain the lower dimensional representation of the data in a time window. Let the real $r \times c$ matrix $\mathbf{M}$ denote the weather data in the time interval before the current time. c is the length of the time window and r is the number of weather variables. Then, the SVD decomposition of $\mathbf{M}$ can be written as

$$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \tag{1}$$

where $\mathbf{U}$ is an $r \times r$ orthogonal matrix, $\mathbf{V}$ is a $c \times c$ orthogonal matrix, and $\mathbf{\Sigma}$ is an $r \times c$ diagonal matrix with entries $(\sigma_1, \sigma_2, \ldots, \sigma_p)$, where $p = \min(r, c)$ and $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0$. The σ are the singular values of $\mathbf{M}$ and the columns of $\mathbf{U}$ and $\mathbf{V}$ are the left and right singular vectors, respectively, of $\mathbf{M}$. If we consider only the top largest k singular values (also called the principal singular values) and the corresponding column vectors of $\mathbf{U}$ and $\mathbf{V}$, which form the sub-matrices $\mathbf{U}_k$ and $\mathbf{V}_k$, respectively, we obtain a reduced dimensional representation $\mathbf{M}_k$ of $\mathbf{M}$ as follows:

$$\mathbf{M}_k = \mathbf{U}_k\mathbf{\Sigma}_k\mathbf{V}_k^T. \tag{2}$$

In signal processing, this reduced dimensional space is often called the signal space as k is chosen so that the singular values that are excluded are small and form the noise in the data.

Let $\mathbf{M}_{old}$ denote the matrix $\mathbf{M}$ at the time window that ends at time t, that is, $\mathbf{M}_{old}$ contains data from time steps $t-c+1$, $t-c+2$, …, t in order. Suppose we choose to keep k singular values such that the reduced representation captures $P\%$ of the energy in the data. We expect that both k and the corresponding singular values in some way represent the matrix $\mathbf{M}_{old}$. When new data at time $(t+1)$ arrive, the sliding window shifts by one and $\mathbf{M}_{new}$ now contains data from time steps $t-c+2$, $t-c+3$, …, t, $t+1$ in order.

Our goal is to understand if the changes in the singular values and vectors of the signal subspace, as the matrix of observations transitions from $\mathbf{M}_{old}$ to $\mathbf{M}_{new}$, can provide some insight into the wind generation in the time steps just after the current time.

IV. INCREMENTAL SINGULAR VALUE DECOMPOSITION

A naïve approach to implementing SVD on sliding windows is to recalculate the decomposition whenever new data are added and the oldest data are removed from the window. However, the operation count to obtain an SVD of an $r \times c$ matrix is $O(rc^2)$, assuming that $r \gg c$. Instead, we use an approximation that keeps just the information in the lower dimensional signal space, discarding the noise component. As new data arrive, this information in the signal sub-space is updated appropriately, along with any changes in the rank of the signal sub-space.

An elegant formulation of an incremental SVD method, called the Fast Approximate Subspace Tracking (FAST) algorithm, was proposed by Real, Tufts and Cooley in 1997 [4], [5]. We briefly summarize their approach below.

At the time instant t, the matrix $\mathbf{M}_{old}$ can be considered to be the sum of a reduced-rank signal matrix $\mathbf{S}_{old}$ of rank k and a full-rank noise matrix $\mathbf{N}_{old}$:

$$\mathbf{M}_{old} = \mathbf{S}_{old} + \mathbf{N}_{old}$$

Recall that the rows of $\mathbf{M}_{old}$ represent the weather variables and the columns represent the time instants in the window.

The matrices $\mathbf{M}_{old}$ and $\mathbf{M}_{new}$ can be represented in terms of their column vectors $\mathbf{m}_i$ as

$$\mathbf{M}_{old} = [\mathbf{m}_1 \quad \mathbf{m}_2 \quad \ldots \quad \mathbf{m}_c]$$

and

$$\mathbf{M}_{new} = \left[\mathbf{m}_2 \quad \mathbf{m}_3 \quad \ldots \quad \mathbf{m}_{(c+1)}\right]$$

where the oldest values are removed from one end of the matrix and the new ones added at the other end.

Suppose that we have a sufficiently accurate approximation to the k principal singular values and the corresponding left singular vectors of $\mathbf{M}_{old}$. Let these k orthonormal approximate left singular vectors be represented as the columns of an $r \times k$ matrix $\mathbf{U}_{old}$ as

$$\mathbf{U}_{old} = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \ldots \quad \mathbf{u}_k]$$

where $\mathbf{u}_i$ is associated with the i-th largest approximate singular value. These column vectors form the basis for the signal space corresponding to $\mathbf{M}_{old}$. By definition, the error in reconstruction resulting from the use of only the largest k singular values and their corresponding vectors is given by the squared Frobenius norm of the difference between the original matrix $\mathbf{M}_{old}$ and its projection onto the reduced dimension space spanned by the columns of $\mathbf{U}_{old}$:

$$\epsilon_{old} = \|\mathbf{M}_{old} - \mathbf{U}_{old}\mathbf{U}_{old}^T\mathbf{M}_{old}\|_F^2 \ .$$

The FAST algorithm updates the approximate singular value decomposition in two steps. First, it creates a low-rank approximation $\mathbf{A}$, of r rows and c columns, to $\mathbf{M}_{new}$ such that

$$\|\mathbf{M}_{new} - \mathbf{A}\|_F^2 \leq \epsilon_{old}.$$

Thus, if the error ϵ_{old} was acceptable in the prior step, the error in the new approximation will be no greater, and therefore, should be acceptable as well. The second step uses the information in the matrix $\mathbf{A}$ to construct a smaller matrix $\mathbf{F}$, which is then used to obtain the approximate singular values and vectors of $\mathbf{M}_{new}$.

Let the approximation to $\mathbf{M}_{old}$ be written as:

$$\begin{aligned}\mathbf{M}_{old} &\approx \mathbf{U}_{old}\mathbf{U}_{old}^T\mathbf{M}_{old} \\ &= \mathbf{U}_{old}\,[\mathbf{a}_1 \quad \mathbf{a}_2 \quad \ldots \quad \mathbf{a}_c] \\ &= [\mathbf{g}_1 \quad \mathbf{g}_2 \quad \ldots \quad \mathbf{g}_c]\end{aligned}$$

where $\mathbf{a}_j = \mathbf{U}_{old}^T\mathbf{m}_j$ is a $k \times 1$ column vector, $\mathbf{m}_j$ is the j-th column of $\mathbf{M}_{old}$, and $\mathbf{g}_j$ is the j-th column of $\mathbf{U}_{old}\mathbf{U}_{old}^T\mathbf{M}_{old}$. Since $\mathbf{M}_{new}$ differs from $\mathbf{M}_{old}$ in two columns, we can exploit the existing decomposition of the approximation to $\mathbf{M}_{old}$ to create the matrix $\mathbf{A}$, which is the rank $(k+1)$ approximation to $\mathbf{M}_{new}$, as follows:

$$\begin{aligned}\mathbf{A} &= [\mathbf{U}_{old} \quad \mathbf{q}] \begin{bmatrix}\mathbf{a}_2 & \mathbf{a}_3 & \ldots & \mathbf{a}_c & \mathbf{a}_{(c+1)} \\ 0 & 0 & \ldots & 0 & b\end{bmatrix} \qquad (3) \\ &= [\mathbf{U}_{old} \quad \mathbf{q}]\,\mathbf{E}.\end{aligned}$$

The first matrix on the right hand side is an $r \times (k+1)$ matrix, while the second matrix, $\mathbf{E}$, is a $(k+1) \times c$ matrix. The $r \times 1$ column vector $\mathbf{q}$ and the scalar b are obtained by decomposing the new column $\mathbf{m}_{(c+1)}$ into two components - one (that is, $\mathbf{U}_{old}\mathbf{a}_{(c+1)}$), which is in the column space of $\mathbf{U}_{old}$ and one (that is, $b\mathbf{q}$), which is in the space orthogonal to $\mathbf{U}_{old}$:

$$\begin{aligned}\mathbf{a}_{(c+1)} &= \mathbf{U}_{old}^T\mathbf{m}_{(c+1)} \\ \mathbf{z} &= \mathbf{m}_{(c+1)} - \mathbf{U}_{old}\mathbf{a}_{(c+1)} \\ b &= \|\mathbf{z}\| \\ \mathbf{q} &= \frac{\mathbf{z}}{b}\end{aligned}$$

By expanding the right hand side of Equation(3) we obtain

$$\mathbf{A} = \left[\mathbf{g}_2 \quad \mathbf{g}_3 \quad \ldots \quad \mathbf{g}_c \quad \mathbf{m}_{(c+1)}\right].$$

The resulting error in approximating $\mathbf{M}_{new}$ by $\mathbf{A}$ is given by:

$$\|\mathbf{M}_{new} - \mathbf{A}\|_F^2 = \sum_{i=2}^{c}\left\{\|\mathbf{m}_i - \mathbf{g}_i\|^2\right\} + \|\mathbf{m}_{(c+1)} - \mathbf{m}_{(c+1)}\|^2$$

with the second term on the right hand side being zero. This is smaller than the error at the prior step:

$$\begin{aligned}\epsilon_{old} &= \|\mathbf{M}_{old} - \mathbf{U}_{old}\mathbf{U}_{old}^T\mathbf{M}_{old}\|_F^2 \\ &= \sum_{i=1}^{c}\|\mathbf{m}_i - \mathbf{g}_i\|^2,\end{aligned}$$

making $\mathbf{A}$ an acceptable rank $(k+1)$ approximation to $\mathbf{M}_{new}$.

However, $\mathbf{A}$ is the same size as $\mathbf{M}_{new}$. To reduce the amount of computations required to update $\mathbf{U}_{old}$, we want to work with a matrix of a smaller size than $\mathbf{M}_{new}$. We observe that the first matrix on the right hand side of Equation (3) has, by definition, $(k+1)$ orthonormal columns. So, if we construct the singular value decomposition of the second matrix, $\mathbf{E}$, of size $(k+1) \times c$, as follows:

$$\mathbf{E} = \mathbf{U}_E\mathbf{\Sigma}_E\mathbf{V}_E^T,$$

we can generate the singular value decomposition of $\mathbf{A}$ as:

$$\begin{aligned}\mathbf{A} &= \Big([\mathbf{U}_{old} \quad \mathbf{q}]\,\mathbf{U}_E\Big)\mathbf{\Sigma}_E\mathbf{V}_E^T \\ &= \mathbf{U}_A\mathbf{\Sigma}_A\mathbf{V}_A^T\end{aligned}$$

where

$$\begin{aligned}\mathbf{U}_A &= [\mathbf{U}_{old} \quad \mathbf{q}]\,\mathbf{U}_E \qquad (4) \\ \mathbf{\Sigma}_A &= \mathbf{\Sigma}_E \\ \mathbf{V}_A &= \mathbf{V}_E.\end{aligned}$$

This allows us to obtain the $(k+1)$ principal left singular vectors of $\mathbf{A}$ as the columns of the $r \times (k+1)$ matrix $\mathbf{U}_A$ by calculating the singular value decomposition of the smaller matrix $\mathbf{E}$. Since $\mathbf{A}$ is an approximation to $\mathbf{M}_{new}$, this gives us the approximation to the left singular values of $\mathbf{M}_{new}$. The approximation to the singular values of $\mathbf{M}_{new}$ are obtained by considering the $(k+1)$ elements on the main diagonal of $\mathbf{\Sigma}_E$.

It is possible to reduce the computations even further by considering the matrix $\mathbf{F}$ defined as:

$$\begin{aligned}\mathbf{F} &= \mathbf{E}\mathbf{E}^T \qquad (5) \\ &= \big(\mathbf{U}_E\mathbf{\Sigma}_E\mathbf{V}_E^T\big)\big(\mathbf{V}_E\mathbf{\Sigma}_E\mathbf{U}_E^T\big) \\ &= \mathbf{U}_E\mathbf{\Sigma}_E\mathbf{\Sigma}_E\mathbf{U}_E^T \qquad (6) \\ &= \mathbf{U}_F\mathbf{\Sigma}_F\mathbf{V}_F^T.\end{aligned}$$

The matrix $\mathbf{F}$ is a smaller $(k+1) \times (k+1)$ matrix and its singular value decomposition can be obtained more easily than that of $\mathbf{E}$, which is a $(k+1) \times c$ matrix.

We can now obtain the singular values and vectors of $\mathbf{A}$ as follows. The singular vectors of $\mathbf{F}$ are the columns of $\mathbf{U}_F$. Since these are also the columns of $\mathbf{U}_E$, we can use Equation(4) to calculate the left singular vectors of $\mathbf{A}$. In addition, from Equation(6), if follows that the singular values of $\mathbf{A}$ are the square-root of the singular values of $\mathbf{F}$.

In summary, the FAST algorithm creates an approximation to the SVD of $\mathbf{M}_{new}$ by calculating the singular values and vectors of $\mathbf{F}$ and combining them with $\mathbf{U}_{old}$ from the previous step and the vector $\mathbf{q}$, which is obtained using $\mathbf{U}_{old}$ and the vector $\mathbf{m}_{(c+1)}$ representing the new values in $\mathbf{M}_{new}$. The steps in the FAST method are summarized in Algorithm 1 [6].

Algorithm 1 *FAST algorithm.*

Obtain initial estimate of k principal singular values and left singular vectors $\mathbf{U}_{old}$ of the initial matrix $\mathbf{M}$
while new data arrive **do**
  Obtain new data vector $\mathbf{m}_{(c+1)}$
  $\mathbf{a}_i = \mathbf{U}_{old}^T \mathbf{m}_i, i = 2, 3, \ldots, (c+1)$
  $\mathbf{z} = \mathbf{m}_{(c+1)} - \mathbf{U}_{old}\mathbf{a}_{(c+1)}$
  $b = \|\mathbf{z}\|$
  $\mathbf{q} = \mathbf{z}/b$
  $\mathbf{E} = \begin{bmatrix} \mathbf{a}_2 & \mathbf{a}_3 & \ldots & \mathbf{a}_c & \mathbf{a}_{(c+1)} \\ 0 & 0 & \ldots & 0 & b \end{bmatrix}$
  $\mathbf{F} = \mathbf{E}\mathbf{E}^T$
  Compute SVD: $\mathbf{F} = \mathbf{U}_F \mathbf{\Sigma}_F \mathbf{V}_F^T$
  Replace columns of $\mathbf{U}_{old}$ with columns of $[\mathbf{U}_{old} \quad \mathbf{q}]\,\mathbf{U}_F$
  Replace old singular values with square root of singular values of $\mathbf{\Sigma}_F$
  Update data vectors: $\mathbf{m}_i \leftarrow \mathbf{m}_{(i+1)}, i = 1, 2, \ldots, c$
end while

The FAST algorithm can also be used to detect any changes in the dimension of the signal subspace as the matrix of observations changes from $\mathbf{M}_{old}$ to $\mathbf{M}_{new}$ ([5] Section VIII). This dimension is obtained by considering the top k singular values and vectors such that a reconstruction using them explains $P\%$ of the energy in the data, which is the sum of the squares of the singular values of the matrix or the the square of the Frobenius norm of the matrix. Given the Frobenius norm of $\mathbf{M}_{old}$, the norm of $\mathbf{M}_{new}$ is obtained as:

$$\|\mathbf{M}_{new}\|_F^2 = \|\mathbf{M}_{old}\|_F^2 - \|\mathbf{m}_1\|_2^2 + \|\mathbf{m}_{(c+1)}\|_2^2$$

Suppose $\mathbf{M}_{old}$ was represented using k approximate singular values and vectors. If the energy calculation for $\mathbf{M}_{new}$ indicates that the signal subspace dimension has increased, we keep all the $(k+1)$ approximate singular values and vectors calculated using Algorithm 1. However, if the signal subspace dimension has reduced or remained at k, the corresponding number of approximate singular values and vectors can be retained at the end of the processing of $\mathbf{M}_{new}$.

The FAST algorithm stores the data for the previous decomposition and thus requires a memory of at least rc to store $\mathbf{U}_{old}$, where r is the dimension of the original data. Its computational cost is $O(rk^2)$ [5].

V. Experimental results

In our analysis using data from the mid-Columbia Basin region, we consider three window sizes - $c = 12, 24$, and 36 - representing 60, 120 and 180 minutes (1, 2 and 3 hours), respectively. We expect that a very small window size is unlikely to provide enough information on the generation in the future, and too large a window size would result in poor accuracy of prediction as it would capture outdated weather information. We also set P, the percentage of energy in the data to be retained as 99%. In the current analysis, we focus on the number of singular values, denoted by k, that capture $P\%$ of the energy, as well as the largest singular value, denoted by s, as these appeared to be most relevant to predicting the generation in the time period just after the end of the window.

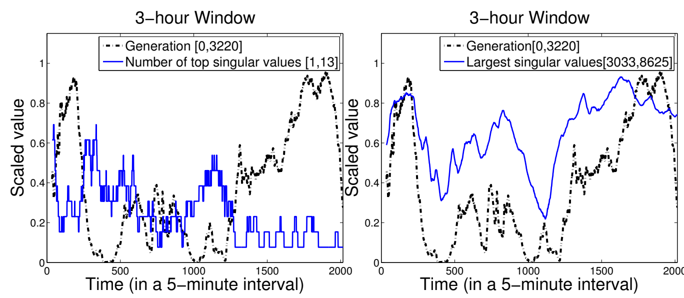


Fig. 3. The number of singular values (left) and the largest singular values (right) obtained from incremental SVD using a three-hour window ($c = 36$) for the first week in October 2011.

Figure 3 shows how s and k vary with time for the first week in October 2011, along with the wind generation. These results are obtained by applying the FAST algorithm to weather data using a three-hour ($c = 36$) sliding window. The wind generation, s, and k for the entire month of October 2011 have each been scaled to lie between [0,1], with the figure showing just the first week. The ranges of s, k, and generation for the month are [3033,8625], [1,13], and $[0, 3220]$ MW, respectively.

We observe that there is an inverse relationship between k and the wind generation. When the generation is high, the number of singular values necessary to represent 99% of the energy in the data is small, and vice-versa. In contrast, the largest singular value, s, follows the generation and tends to be high when the generation is high. This indicates that s and k are likely to be predictive of the wind generation.

To explore this idea further, we consider how the distribution of s and k, obtained for each time point, t, in October 2011, using the weather data in the time window $[t-c+1, t]$, relates to the magnitude of the change in the generation, $|\Delta W|$, during the time period $[t, t+l]$. In other words, do s and k for the time period just before t, reflect the change in generation in the time period just after t? Our results are presented in Figure 4, where we plot the largest singular values, s, on the x-axis and the number of top singular values, k, on the y-axis. Each point represents a time instant t, with the points separated based on whether $|\Delta W|$ is high ($>$ 100W) or low ($\leq$ 100W). We show two values of l: 10 mins and 30 mins. The colors indicate increase and decrease in generation.

It is clear from Figure 4 that $|\Delta W|$ is high only when the largest singular values are > 5000 and the number of top singular values are < 8. Or, $|\Delta W|$ is likely to be low if the

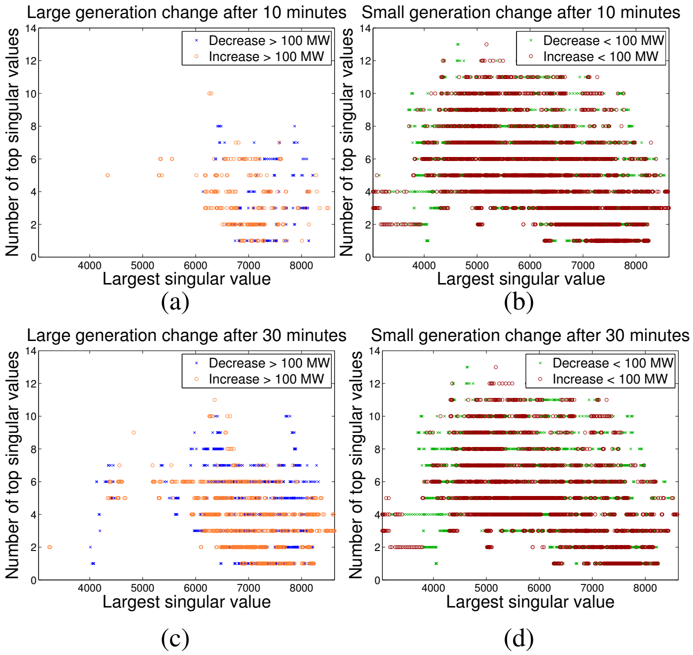


Fig. 4. The largest singular values, s, on the x-axis, and the numbers of top singular values, k, on the y-axis, obtained using a three-hour window ($c = 36$) for October 2011. (a) and (c) show time points for which the generation change $|\Delta W| > 100$ MW, while (b) and (d) show time points for which $|\Delta W| \leq 100$ MW. The time interval l for the generation change is 10 mins for (a) and (b) and 30 mins for (c) and (d). The colors indicate increase and decrease in generation.

values of k and s are not in the lower right corner of the plot. For a larger l, the range of s and k for high $|\Delta W|$ is larger.

This observation allows us to create a simple decision boundary based on the values of s and k. Since we care more about larger changes in generation, we compute the sensitivity as the proportion of correctly predicted high generation changes and the specificity as the proportion of correctly predicted low generation changes. The overall accuracy considers the proportion of all correctly predicted time points.

To determine the decision boundary for high $|\Delta W|$, we consider the distribution of s and k for $|\Delta W| > 100$ and $l = 1$, and choose the thresholds such that all points are within the decision boundary, that is, the specificity is 100%. We found that the values of s and k that determine the decision boundary change slightly with the length of the sliding window $c = 12, 24,$ and 36 as shown in Table II.

Window length c	**Ranges with high $\|\Delta W\|$**
12	$s > 3000$ and $k < 7$
24	$s > 3000$ and $k < 9$
36	$s > 5000$ and $k < 8$

TABLE II. THRESHOLDS USED FOR GROUPING THE HIGH GENERATION CHANGES, $|\Delta W| > 100$ MW, FOR $c = 12, 24$ AND 36 FOR OCTOBER 2011 DATA.

Figure 5(a) and (b) shows the sensitivity and specificity, respectively, in prediction of $|\Delta W|$ for October 2011 for the three time windows and different values of $l = 1, 2, 3, \ldots$, corresponding to $5, 10, 15, \ldots$ minutes after the current time. We observe that all time windows used in incremental SVD behave similarly using the thresholds in Table II. Consider the

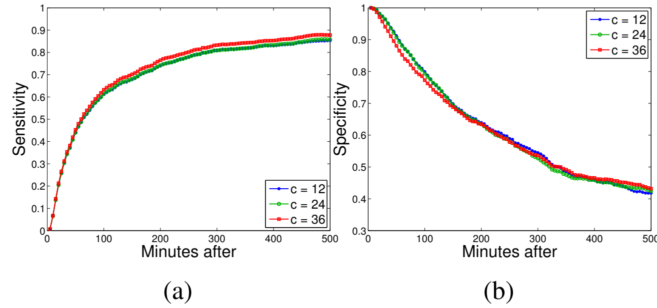


Fig. 5. (a) Sensitivity and (b) specificity in $|\Delta W|$ after 5,10, 15, ... minutes ($l = 1, 2, 3, \ldots$) in October 2011, using the thresholds in Table II. The results are presented for $c = 12, 24$ and 36.

results for $l = 6$, that is the generation change 30 mins after current time, as shown in the lower plots in Figure 4. The incremental SVD with $c = 36$ gives only 31.2% sensitivity on the proportion of high $|\Delta W|$ using the thresholds $s > 5000$ and $k < 8$. For low $|\Delta W|$, it gives a specificity that reaches up to 94.1%. Although the overall accuracy (see Figure 6(a)) is only 48.1% for the prediction after 30 minutes, we can still be very confident on the low generation changes when the values s and k are outside the thresholds.

We also observe from Figure 5(a) that as time increases, there are more high generation changes clustered in the region defined in Table II. When s and k values fall in this region, 70% of the time $|\Delta W| > 100$ MW within three hours (180 mins). Figure 5(b) shows that when s and k values fall outside the ranges with high $|\Delta W|$, at least 90% of the time the generation change in the next 50 minutes will be low.

Thus far, we have considered only the weather and wind generation data from the month of October 2011. We next consider how the decision thresholds derived in Table II will perform when used in prediction of $|\Delta W|$ on test data from November 2011. Figure 6(a) and (b) show the overall accuracy for the October 2011 training data and the November 2011 testing data, respectively. The overall accuracy of the incremental SVD with $c = 12$ and 36 on predicting $|\Delta W|$ reaches above 70% within an hour in November.

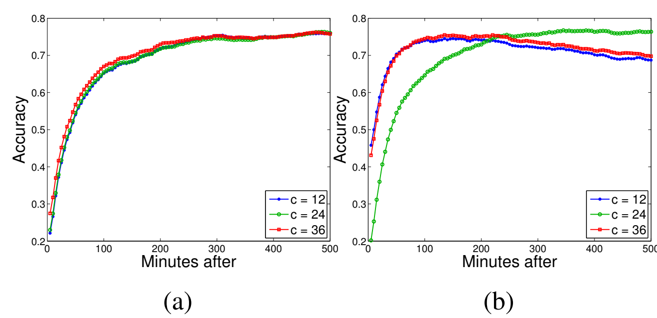
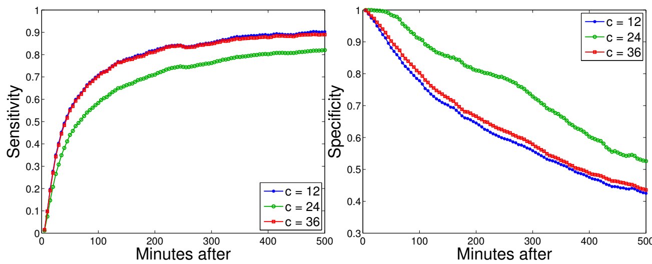


Fig. 6. Overall accuracy in training (a) and testing (b) for $|\Delta W|$ after 5,10, 15, ... minutes ($l = 1, 2, 3, \ldots$), using the thresholds in Table II. The results are presented for $c = 12, 24$ and 36.

Figure 7 displays the sensitivity and specificity on the test data. When $c = 12$ and 36, the sensitivity and specificity are very similar to the October 2011 training data. When s and k fall outside the ranges associated with high $|\Delta W|$ in the training data, we can say that for the test data, more than 90% of the time, there will be no $|\Delta W| > 100$ MW within 50 minutes. Similarly, when s and k fall within the ranges, 70%

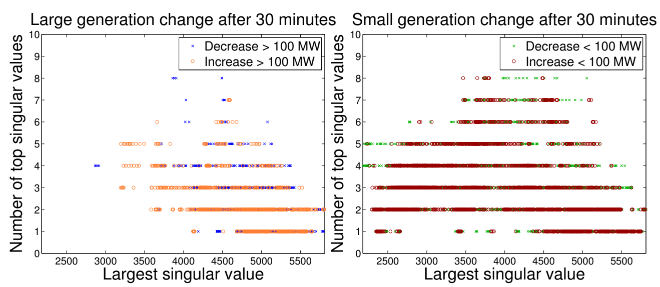
of the time we expect to see $|\Delta W| > 100$ within 2 hours.



(a) (b)

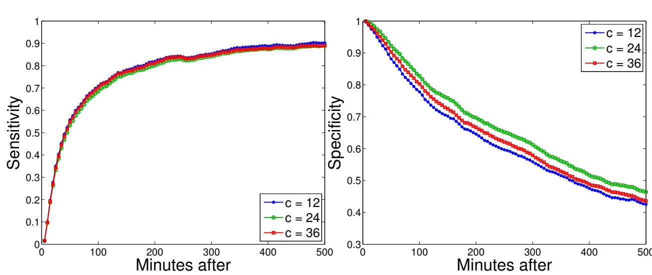
Fig. 7. Sensitivity and specificity in predicting $|\Delta W|$ after 5,10, 15, ... minutes ($l = 1, 2, 3, \ldots$) in November 2011, using the thresholds in Table II obtained from training on October 2011 data. The results are presented for $c = 12, 24$ and 36.

However, for $c = 24$, the accuracy in testing, the sensitivity, and the specificity are all different from the corresponding curves for $c = 12$ and 36. This suggests that the thresholds used from the October 2011 data may not be the most appropriate for this time window. To confirm this, we plot the distribution of s and k for the high and low generation change for November 2011 as shown in Figure 8. This indicates a revised decision boundary of $s > 3800$ and $k < 7$, which results in $c = 24$ giving similar performance to $c = 12$ and 36 as shown in Figure 9.



(a) (b)

Fig. 8. The largest singular values, s, on the x-axis, and the numbers of top singular values k, on the y-axis, obtained using a two-hour window ($c = 24$) for November 2011. (a) and (b) show time points for which $|\Delta W| > 100$ MW and $|\Delta W| \leq 100$ MW, respectively. The time interval $l = 6$ (= 30 mins). The colors indicate increase and decrease in generation.



(a) (b)

Fig. 9. Sensitivity and specificity in predicting $|\Delta W|$after 5,10, 15, ... minutes ($l = 1, 2, 3, \ldots$) in November 2011, using the **updated thresholds**, $s > 3800$ and $k < 7$ for $c = 24$. The results are presented for $c = 12, 24$ and 36.

This suggests that we need to select the thresholds for the decision boundaries more carefully. In future work, we will explore clustering and classification methods, as well as an improved set of variables (in addition to just s and k) to identify more robust decision boundaries.

VI. Related Work

Many ideas have been explored to improve the predictive power of wind generation forecasts [7] and to use data mining techniques to support the scheduling of wind generation [8], [1], [2]. To the best of our knowledge, there has not been any work in exploiting a representation derived from a singular value decomposition of the weather data in the time period just before the current time to predict the wind generation in the near future. In our analysis, we have used one approach [5] to calculate the singular value decomposition incrementally. Other approaches [9], [10], [11] are also possible, though we expect that they would give similar results.

VII. Conclusions

In this paper, we show how we can use the SVD of weather data in a time window just before the current time to predict the change in wind generation in the time interval just after the current time. Our approach is very effective in predicting small changes in generation and computationally efficient as it uses an incremental algorithm to analyze the weather data streams as they arrive.

Acknowledgment

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

References

[1] C. Kamath and Y. J. Fan, "Finding motifs in wind generation time series data," in *Machine Learning and Applications (ICMLA), 2012 11th International Conference on*, vol. 2, 2012, pp. 481–486.

[2] ——, "Using data mining to enable integration of wind resources on the power grid," *Statistical Analysis and Data Mining*, vol. 5, no. 5, pp. 410–427, 2012.

[3] "Bonneville Power Administration Wind Power web page," http://www.bpa.gov/corporate/windpower/, August 2012.

[4] D. W. Tufts, E. C. Real, and J. Cooley, "Fast approximate subspace tracking (FAST)," in *Proceedings, IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 1, 1997.

[5] E. C. Real, D. W. Tufts, and J. Cooley, "Two algorithms for fast approximate subspace tracking," *IEEE Transactions on Signal Processing*, vol. 47, no. 7, pp. 1936–1945, July 1999.

[6] C. Kamath, "Dimension reduction for streaming data," in *Data-Intensive Computing: Architectures, Algorithms, and Applications*, I. Gorton and D. K. Gracio, Eds. Cambridge University Press, 2013, pp. 124–156.

[7] C. Monteiro *et al.*, "Wind power forecasting: State-of-the-art 2009," Argonne National Laboratory, Tech. Rep., November 2009.

[8] H. Zareipour, D. Huang, and W. D. Rosehart, "Wind power ramp events classification and forecasting: A data mining approach," in *Proceedings, IEEE PES Annual General Meeting*, July 2011.

[9] P. Strobach, "Low rank adaptive filters," *IEEE Trans. on Signal Processing*, vol. 44, no. 12, pp. 2932–2947, 1996.

[10] M. Brand, "Fast online SVD revisions for lightweight recommender systems," in *Proceedings, SIAM International Conference on Data Mining*, 2003, pp. 37–46.

[11] P. Strobach, "The fast recursive row-Householder subspace tracking algorithm," *Signal Processing*, vol. 89, pp. 2514–2528, 2009.